

Санкт-Петербургский государственный университет  
Фундаментальная информатика и информационные технологии  
Информационные технологии

Бирилло Анастасия Игоревна

Алгебраические байесовские сети:  
представление данных, алгоритмы  
обработки и реинжиниринг комплекса  
программ (проектная работа)

Бакалаврская работа

Научный руководитель:  
профессор кафедры информатики, д. ф.-м. н., доцент Тулупьев А. Л.

Рецензент:  
доцент кафедры компьютерных технологий, к. ф.-м. н. Фильченков А. А.

Санкт-Петербург  
2017

SAINT-PETERSBURG STATE UNIVERSITY

Fundamental Informatics and Information Technology  
Information Technology

Anastasia Birillo

Algebraic Bayesian Networks:  
data representation, processing algorithms  
and programme complex reengineering (joint  
project)

Graduation Project

Scientific supervisor:  
Prof. Computer Science Department, Dc. Sc. in Math, Assoc. Prof. Alexander Tulupyev

Reviewer:  
As. Prof. Computer Technology Department, Ph. D. Sc. in Math Andrey Filchenkov

Saint-Petersburg  
2017

# Оглавление

<b>Введение</b>	<b>4</b>
<b>1. Алгебраические байесовские сети: структура и оценки вероятности истинности</b>	<b>9</b>
1.1. Введение . . . . .	9
1.2. Структура АБС . . . . .	9
1.3. Локальный апприорный и апостериорный вывод . . . . .	13
1.4. Выводы по главе . . . . .	19
<b>2. Алгебраические байесовские сети: база данных</b>	<b>20</b>
2.1. Введение . . . . .	20
2.2. Выбор технологий . . . . .	20
2.3. Схема базы данных . . . . .	24
2.4. Существующие проблемы и их решение . . . . .	28
2.5. Выводы по главе . . . . .	30
<b>3. Расширенная документация:</b>	
<b>материалы для методического пособия</b>	<b>31</b>
3.1. Введение . . . . .	31
3.2. Структура . . . . .	31
3.3. Аналитическая проверка: численные эксперименты . . .	33
3.4. Программная проверка . . . . .	36
3.5. Выводы по главе . . . . .	39
<b>Заключение</b>	<b>40</b>
<b>Список литературы</b>	<b>41</b>

# Введение

**Актуальность темы исследования.** Теория вероятностных графических моделей (ВГМ) [36] является одной из областей искусственного интеллекта, дающей возможность связать понятия данных, вероятностной логики и теории графов. Соединяя эти понятия, исследователь получает возможность анализировать данные, принимая во внимание их связанность между собой (например, последовательность состояний или причинно-следственные связи). К ВГМ относятся такие модели как марковские цепи и байесовские сети доверия, имеющие широкое применение и родственные алгебраическим байесовским сетям, рассматриваемым в данной работе. Вероятностные графические модели имеют широкое применение в ряде областей и давно подтвердили свою состоятельность [5, 9, 10, 14]. Они в настоящее время активно применяются в системах распознавания образов [22, 25], интеллектуальных системах поддержки принятия решений и рекомендательных системах [3, 4, 6, 13], системах моделирования рисков и обнаружения отказов [7, 15, 18, 31], оценивания повреждений и мониторинга состояния здоровья [1, 2, 8] и системах моделирования отклика экологических систем на различные изменения и воздействия [11, 17].

Одним из видов ВГМ является структура алгебраических байесовских сетей (АБС) [26, 27], предполагающая декомпозицию базы знаний на небольшие объемы данных, тесно связанных между собой, что позволяет ускорить алгоритмы обработки за счет возможности работать с ограниченным набором элементов. Иным фактором, влияющим на качество и скорость работы алгоритмов модели, является постановка решаемых задач. Кроме того, АБС предполагают наличие интервальных вероятностных оценок, что дает возможность расширить оценки на узлах фрагментов знаний.

В настоящее время ведется работа над проектом web-платформы «AlgBN Web App» по работе и обучению АБС [28]. В рамках данного проекта используется математическая библиотека, основанная на комплексе программ «AlgBN Math Library» [19, 30], который не имеет до-

статочной подробной обучающей документации для быстрого освоения. Так как на данном комплексе программ основана большая часть существующих ныне исследований, то был необходим способ быстрого ознакомления с комплексом для его освоения. В связи с этим в данной работе были созданы кейсы тестов, охватывающие всю функциональность комплекса программ, которые были оформлены в расширенную документацию по комплексу программ, что является основой для создания методического пособия по основам теории АБС и комплексу программ для их обработки.

Также разработанный комплекс программ не включает в себя базу данных, позволяющую удаленно хранить данные для работы с ними. Существующее решение [33] (разработанная схема базы данных) является устаревшим, не учитывающим новейшие теоретические результаты. Кроме того, разрабатываемая web-платформа также нуждается в базе данных для хранения АБС и работы с ними. Поэтому в рамках данной работы была разработана и реализована база данных для хранения АБС и работы с АБС.

**Целью данной выпускной квалификационной работы** является создание базы данных для хранения АБС и работы с ними с сопутствующей подготовкой материалов для методического пособия в рамках web-платформы «AlgBN Web App» по обучению теории АБС. Для достижения данной цели были поставлены **следующие задачи**:

1. Проектирование схемы базы данных;
2. Реализация базы данных средствами СУБД MS SQL Server;
3. Создание уровня доступа к данным (DAL);
4. Разработка кейсов тестов к комплексу программ «AlgBN Math Library» и их реализация на языке C#;
5. Создание расширенной документации к комплексу программ «AlgBN Math Library».

**Методы исследования.** Для решения поставленных задач в данной области понадобилось изучить методические материалы, поставить проблему, проанализировать ее, спроектировать несколько вариантов решения. Затем были выбраны подходящие средства и технологии программирования, связанные с языком реализации (C#, SQL), средой разработки (Visual Studio, MS SQL Sever). После чего были проведены численные и программные эксперименты, позволяющие написать методическое пособие. А также были проведены тестовые случаи для обоснования корректности полученной базы данных. В качестве методов решения используются теоретические методы (анализ предыдущих результатов), методы объектно-ориентированного программирования (ООП), методы проектирования баз данных.

**Публикации.** По теме дипломного проекта было подготовлено 6 научных публикаций:

- А.И. Бирилло Локальный синтез согласованных оценок истинности в алгебраических байесовских сетях: вычислительные эксперименты // Материалы 6-й всероссийской научной конференции по проблемам информатики СПИСОК-2016. (26–29 апреля 2016 г. Санкт-Петербург). СПб.: ВВМ, 2016. С. 453–459.
- А.И. Бирилло, Анализ согласованности оценок истинности в локальном априорном и апостериорном выводе в алгебраических байесовских сетях: вычислительные эксперименты // Региональная информатика (РИ-2016). XV Санкт-Петербургская международная конференция «Региональная информатика (РИ-2016)» (Санкт-Петербург, 26-28 октября 2016 г.): Материалы конференции. СПб.: СПОИСУ, 2016, С. 515
- А.А. Золотин, Е.А. Мальчевская, А.И. Бирилло, А.Л. Тулупьев, Управление согласованностью оценок вероятностей в локальном апостериорном выводе в алгебраических байесовских сетях // 9-я Российская мультikonференция по проблемам управления, материалы 9-й конференции "Информационные технологии в управле-

нии” (ИТУ-2016) - СПб: АО ”Концерн ”ЦНИИ ”Электроприбор”, 2016, С. 52 – 61.

- А. А. Золотин, Д. Г. Левенец., М. А. Зотов, А. И. Бирилло, А. И. Березин, А. В. Иванова, А. Л. Тулупьев, Алгоритмы обработки и визуализации алгебраических байесовских сетей. // Образовательные технологии и общество, 2017.
- E. Malchevskaya, N. Kharitonov, A. Zolotin, A. Birillo, Algebraic Bayesian Networks: Probabilistic-Logic Inference Algorithms and Storage Structures //Finnish-Russian University Cooperation in Telecommunications, 3 - 7 april 2017 (в печати)
- E. Malchevskaya, A. Birillo, A. Zolotin, A. Tulupyev, Local Propagation Algorithms in Algebraic Bayesian Networks and Related Models: Final Rendering to Matrix-Vector Equations // Intelligent Information Technologies for Industry, 14 - 16 september 2017 (на рецензировании)

Кроме того была отправлена заявка на регистрацию в Роспатент. Данная выпускная квалификационная работа бакалавра содержит материалы исследований, частично поддержанных грантом РФФИ 15-01-09001 — «Комбинированный логико-вероятностный графический подход к представлению и обработке систем знаний с неопределенностью: алгебраические байесовские сети и родственные модели».

**Структура и объем работы.** Работа состоит из введения, трех глав, заключения и используемой литературы.

Первая глава носит обзорный характер и состоит из 4 разделов. В первом разделе содержатся общие сведения об АБС. Во втором разделе подробно описана структура АБС, а также даны все необходимые определения и понятия по данной тематике. В третьем разделе изложены основные теоретические сведения по локальному апприорному и апостериорному выводу в АБС. Четвертый раздел содержит основные выводы по данной главе.

Вторая глава описывает разработанную базу данных и состоит из 5 разделов. Первый раздел описывает существующее решение и структуру данной главы. Второй раздел содержит обзор на существующие технологии и обоснование выбранной СУБД. Третий раздел содержит подробное описание разработанной базы данных с соответствующими схемами. Четвертый раздел описывает существующие при проектировании и реализации проблемы, а также их решение. Пятый раздел содержит выводы по текущей главе.

Третья глава содержит описание созданных материалов к методическому пособию к комплексу программ «AlgBN Math Library» и состоит также из 5 разделов. Первый раздел описывает необходимость создания данного пособия, а также используемые при создании технологии. Второй раздел содержит подробное описание структуры разработанных материалов (основные разделы, их описание). Третий раздел описывает аналитическую проверку разработанных кейсов тестов. Четвертый раздел содержит подробное описание программной проверки для кейсов тестов. Пятый раздел содержит выводы по данной главе.



# 1. Алгебраические байесовские сети: структура и оценки вероятности истинности

## 1.1. Введение

В настоящее время существует множество математических моделей для представления знаний с неопределенностью. Алгебраическая байесовская сеть является одним из особых подклассов вероятностных графических моделей. Знания, лежащие в основе таких моделей, формируются экспертами в предметной области и дают базу для построения дальнейших суждений и получения недостающих знаний. На основании базы знаний строится система утверждений, где моделью утверждения является пропозициональная формула. Особенностью такой структуры является возможность задания оценок вероятностей утверждениям, при этом оценки вероятности могут быть как скалярные, так и интервальные.

При рассуждении о некотором объекте или области, мы опираемся на знания, полученные ранее и, исходя из этой информации, принимаем какие-либо решения и делаем какие-либо выводы. Неопределенность знаний в модели является следствием того, что каждому утверждению, а также ассоциированной с ним пропозициональной формуле, приписана некоторая оценка вероятности.

На основе ряда источников [29, 32, 33, 34, 35] в данной главе будет введен необходимый математический аппарат, формально описывающий алгебраическую байесовскую сеть как модель знаний с неопределенностью.

## 1.2. Структура АБС

Для объединения набора утверждений в теории алгебраических байесовских сетей используется декомпозиция данных на фрагменты знаний. Фрагмент знаний представляет собой набор тесно связанных между собой утверждений с приписанными им вероятностными оценками.

ми [33]. При это сами фрагменты знаний могут не иметь тесной связи между собой. Эксперты в предметной области обычно задают всевозможные зависимости между парами-тройками атомарных утверждений, именно поэтому для уменьшения количества вычисляемых вероятностей используется разбиение на фрагменты знаний. Фрагмент знаний представляет собой совокупность тесно связанных между собой утверждений с приписанными к ним оценками вероятностей (скалярными или интервальными).

В классическом представлении фрагменты знаний строятся над идеалом конъюнктов [33]. Но в настоящее время существуют и альтернативные способы представления фрагментов знаний — идеал дизъюнктов [35] и множество пропозиций-квантов [29].

Сперва обратим внимание на идеал конъюнктов [33]

### Определение 1.1

$$C = \{x_{i_1}x_{i_2} \cdots x_{i_k} | 0 \leq i_1 < i_2 < \cdots < i_k \leq n-1, 0 \leq k \leq n\},$$

где  $x_{i_1}x_{i_2} \cdots x_{i_k}$  означает конъюнкцию соответствующих переменных. Сам знак конъюнкции для удобства записи опустим.

Идеал представляет собой все возможные конъюнкции атомов заданного алфавита, в том числе пустую конъюнкцию и сами атомы. Литерал  $\tilde{x}_i$  означает, что на его месте в пропозициональной формуле может стоять либо  $x_i$ , либо его отрицание  $\overline{x}_i$ .

**Определение 1.2** Множество пропозиций-квантов над алфавитом  $A = \{x_i\}_{i=0}^{n-1}$  определяется аналогичным образом [29]:

$$Q = \{\tilde{x}_0\tilde{x}_1 \cdots \tilde{x}_{n-1}\}.$$

Таким образом квант представляет собой конъюнкцию, которая для любой переменной из алфавита содержит либо ее саму, либо ее отрицание. С учетом заданного алфавита каждому кванту вида  $\tilde{x}_0\tilde{x}_1 \cdots \tilde{x}_{n-1}$  можно сопоставить характеристический вектор, а этому вектору двоичное число, которое рассматривается как номер кванта (или другими словами индексирует квант).

**Определение 1.3** Идеал дизъюнктов имеет вид [35]:

$$\{x_{i_1} \vee x_{i_2} \vee \dots \vee x_{i_k} | 0 \leq i_1 < i_2 < \dots < i_k \leq n-1, 0 \leq k \leq n\},$$

где  $x_{i_1} \vee x_{i_2} \vee \dots \vee x_{i_k}$  означает дизъюнкцию соответствующих переменных. Идеал дизъюнктов содержит все возможные дизъюнкции атомов заданного алфавита, в том числе пустую дизъюнкцию.

Далее слова математическая модель мы будем опускать, поэтому при использовании слов фрагмент знаний мы будем подразумевать математический термин. В свою очередь сам набор утверждений (идеал конъюнктов, дизъюнктов или множество пропозиций-квантов), над которыми строится фрагмент знаний, называется носителем. При этом функция, задающая оценки вероятности высказываниям, может быть как скалярной, так и интервальной. В теории АБС такие фрагменты знаний носят название «фрагмент знаний над скалярными (интервальными) оценками» соответственно.

**Определение 1.4** Зафиксируем конечное множество атомарных пропозициональных формул — упорядоченный алфавит атомов

$$A = \{x_i\}_{i=0}^{n-1} \text{ [33].}$$

Стоит отметить, что индексация атомов задает порядок в алфавите, который будет использован далее. Иными словами, алфавит атомов представляет собой набор элементарных высказываний о предметной области.

На Рис. 1 представлены примеры всех трех вариантов фрагментов знаний над тремя атомами  $\{x_1, x_2, x_3\} \subset A$ , где  $A$  — алфавит, над которым построен фрагмент знаний.

Для перехода к понятию локального апостериорного вывода в АБС, необходимо ввести понятие непротиворечивости данных [33], так как все операции в АБС производятся над согласованными (непротиворечивыми) данными.

**Определение 1.5** Пусть задан фрагмент знаний со скалярными оценками  $\langle C, P_q \rangle$  (здесь и далее  $C$  обозначает носитель фрагмента знаний

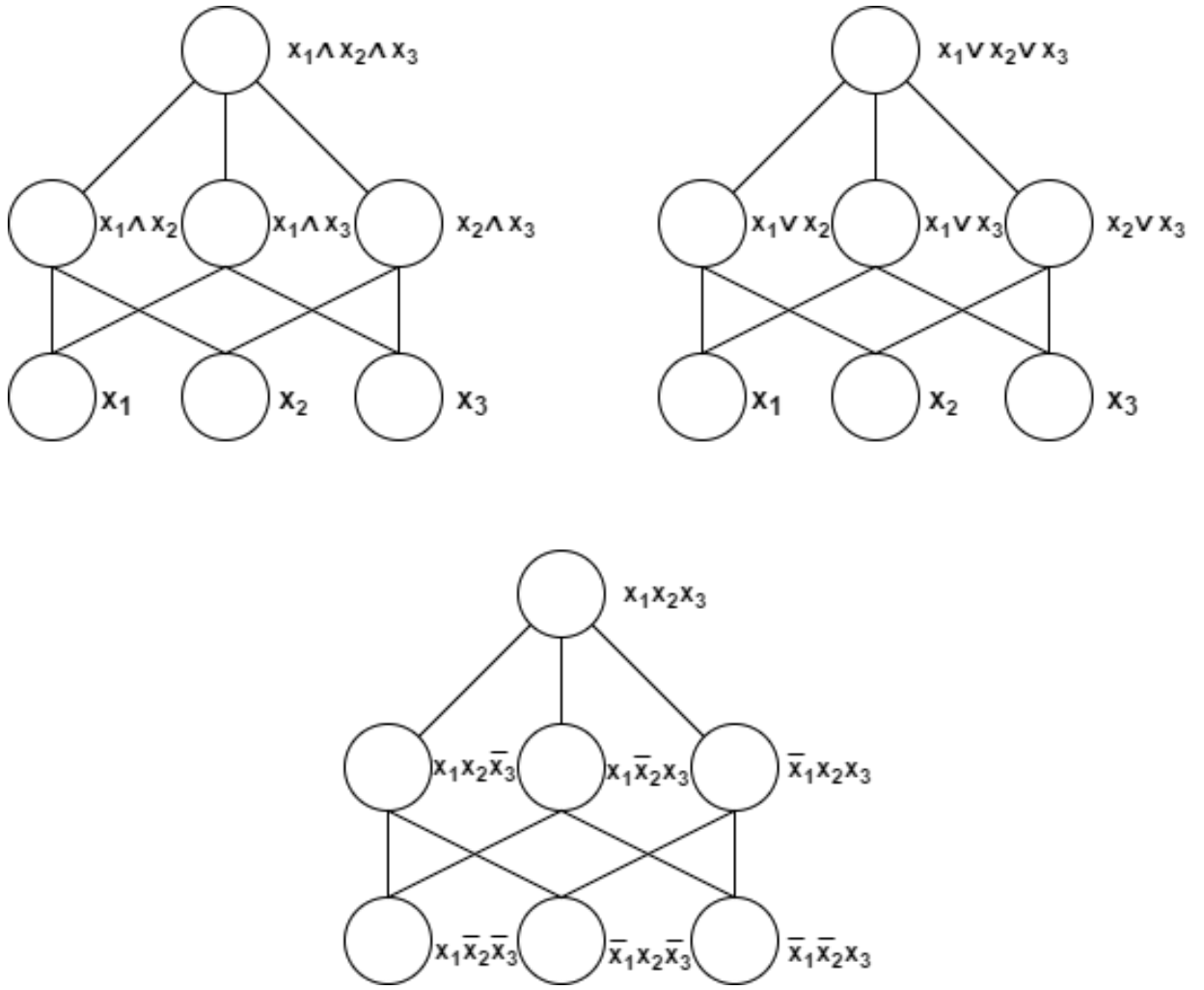


Рис. 1: Фрагмент знаний над алфавитом  $\{x_1, x_2, x_3\}$  над идеалами конъюнктов, дизъюнктов и множеством пропозиций-квантов

— набор утверждений, над которыми он построен). Мы говорим, что он непротиворечив тогда и только тогда, когда существует вероятность  $p_F$ , заданная над множеством пропозициональных формул  $F(A)$ , такая что:

$$\forall c \in C \quad p_F(c) = p(c).$$

$F(A)$  обозначает множество всех пропозициональных формул над алфавитом  $A$  [33].

В случае с интервальным фрагментом знаний необходимо ввести два понятия — согласованный и согласуемый фрагмент знаний [33].

**Определение 1.6** Пусть задан фрагмент знаний с интервальными оценками  $\langle C, \mathbf{P}_q^-, \mathbf{P}_q^+ \rangle$ . Мы говорим, что он непротиворечив (согласо-

ван) тогда и только тогда, когда для любого элемента из носителя фрагмента знаний и любого  $\varepsilon \in p(c)$  найдется функция  $p_{c,\varepsilon} : C \rightarrow [0; 1]$  такая, что  $p_{c,\varepsilon}(c) = \varepsilon$  и  $(C, p_{c,\varepsilon})$  - непротиворечивый фрагмент знаний по определению 1.5.

**Определение 1.7** Пусть задан фрагмент знаний с интервальными оценками  $\langle C, \mathbf{P}_q^-, \mathbf{P}_q^+ \rangle$ . Мы говорим, что он согласуем тогда и только тогда, когда существует непротиворечивый (согласованный) фрагмент знаний с интервальными оценками  $[\mathbf{P}_q^{'-}, \mathbf{P}_q^{' +}]$  такой, что:

$$\forall c \in C \quad [\mathbf{P}_q^{'-}, \mathbf{P}_q^{' +}] \subseteq [\mathbf{P}_q^-, \mathbf{P}_q^+].$$

### 1.3. Локальный априорный и апостериорный вывод

Как было сказано выше, фрагменты знаний могут быть построены над идеалом конъюнктов, дизъюнктов или множеством пропозиций-квантов. Введем обозначение для вектора-вероятностей для каждого случая:

$$\mathbf{P}_c = \begin{pmatrix} 1 \\ p(c_1) \\ \vdots \\ p(c_{2^n-1}) \end{pmatrix}, \quad \mathbf{P}_d = \begin{pmatrix} 0 \\ p(d_1) \\ \vdots \\ p(d_{2^n-1}) \end{pmatrix}, \quad \mathbf{P}_q = \begin{pmatrix} p(q_0) \\ p(q_1) \\ \vdots \\ p(q_{2^n-1}) \end{pmatrix},$$

где  $\mathbf{P}_c$  — обозначение для вектора вероятностей для фрагмента знаний над идеалом конъюнктов;  $\mathbf{P}_d$  — обозначение для вектора вероятностей для фрагмента знаний над идеалом дизъюнктов и  $\mathbf{P}_q$  — обозначение для вектора вероятностей для фрагмента знаний над множеством пропозиций-квантов.

Например, для фрагмента знаний, построенного над двумя атомами  $\{x_1, x_2, x_3\} \subset A$ , изображенного на Рис. 1, векторы вероятностей кван-

тов, конъюнктов и дизъюнктов будут выглядеть следующим образом:

$$\mathbf{P}_c = \begin{pmatrix} 1 \\ p(x_1) \\ p(x_2) \\ p(x_2 x_1) \\ p(x_3) \\ p(x_3 x_1) \\ p(x_3 x_2) \\ p(x_3 x_2 x_1) \end{pmatrix}, \quad \mathbf{P}_q = \begin{pmatrix} p(\bar{x}_3 \bar{x}_2 \bar{x}_1) \\ p(\bar{x}_3 \bar{x}_2 x_1) \\ p(\bar{x}_3 x_2 \bar{x}_1) \\ p(\bar{x}_3 x_2 x_1) \\ p(x_3 \bar{x}_2 \bar{x}_1) \\ p(x_3 \bar{x}_2 x_1) \\ p(x_3 x_2 \bar{x}_1) \\ p(x_3 x_2 x_1) \end{pmatrix}, \quad \mathbf{P}_d = \begin{pmatrix} 0 \\ p(x_1) \\ p(x_2) \\ p(x_2 \vee x_1) \\ p(x_3) \\ p(x_3 \vee x_1) \\ p(x_3 \vee x_2) \\ p(x_3 \vee x_2 \vee x_1) \end{pmatrix}.$$

Заметим, что все три вектора связаны друг с другом (за счет преобразований через вектор квантов). Ниже приведены уравнения, выражающие связь вектора вероятностей конъюнктов и вектора вероятностей квантов [33]:

$$\mathbf{P}_q = \mathbf{I}_n \mathbf{P}_c, \quad \mathbf{P}_c = \mathbf{J}_n \mathbf{P}_q,$$

где  $\mathbf{I}_n = \mathbf{I}_1^{[n]}$ ,  $\mathbf{I}_1 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ , а  $\mathbf{J}_n = \mathbf{J}_1^{[n]}$ ,  $\mathbf{J}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ , где  $[n]$  — степень

Кroneкера (далее эта операция будет обозначаться как  $\otimes$ , что означает тензорное (прямое) произведение матриц). Теперь обратимся к вектору вероятностей дизъюнктов  $\mathbf{P}_d$  и вспомогательному вектору  $\mathbf{P}'_d = \mathbb{1} - \mathbf{P}_d$ , где  $\mathbb{1}$  — единичный вектор такой же длины, как и вектор  $\mathbf{P}_d$ . Для перехода между векторами вероятностей конъюнктов и дизъюнктов будем пользоваться полученными ранее формулами [33, 35]:

$$\mathbf{P}_q = \mathbf{L}_n \mathbf{P}'_d, \quad \mathbf{P}_c = \mathbf{K}_n \mathbf{P}'_d,$$

где  $\mathbf{L}_n = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}^{[n]}$ ,  $\mathbf{K}_n = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}^{[n]}$ .

Как уже упоминалось выше, с точки зрения типов оценок вероятностей АБС могут быть двух видов — скалярные и интервальные. Рассмотрим каждый из типов оценок на примере фрагмента знаний, сформированного над множеством пропозиций-квантов. В случае обработки скалярных оценок вероятностей преобразования будут производиться

с фрагментом знаний  $\langle C, \mathbf{P}_q \rangle$ , а в случае интервальных оценок вероятностей с фрагментом знаний с интервальными оценками  $\langle C, \mathbf{P}_q^-, \mathbf{P}_q^+ \rangle$ . Здесь и далее обозначение  $\mathbf{P}[i]$  подразумевает  $i$ -ю компоненту вектора  $\mathbf{P}$ .

Важным видом логико-вероятностного вывода является задача априорного вывода.

**Определение 1.8** *Априорный вывод заключается в том, чтобы по известным вероятностным оценкам истинности заданных пропозициональных формул построить вероятностную оценку пропозициональной формулы, не вошедшей в число заданных.*

Для описания задачи апостериорного вывода в алгебраических байесовских сетях воспользуемся понятием свидетельства [33]. Под свидетельством понимаются новые «обуславливающие» данные, которые поступают в фрагмент знаний и с учетом которых необходимо пересчитать все (или некоторые) оценки вероятностей. В теории АБС различают три вида свидетельств: детерминированное, стохастическое и неточное. Дадим определение каждому из указанных видов.

**Определение 1.9** *Под детерминированным свидетельством  $\langle i, j \rangle$  понимают предположение о том, что один или несколько атомов получили конкретные означивания. Заметим, что такое свидетельство можно разбить на «положительный» ( $i$ ) и «отрицательный» ( $j$ ) кванты. В первый входят все положительно означенные атомарные переменные свидетельства, а во второй — отрицательно. При этом и положительной, и отрицательной части можно сопоставить индекс, соответствующий десятичному представлению двоичного числа, являющегося характеристическим вектором соответствующей составляющей свидетельства [33].*

**Определение 1.10** *Стохастическое свидетельство — это предположение о том, что над подыдеалом  $C'$  идеала  $C$  задан непротиворечивый фрагмент знаний со скалярными оценками вероятности истинности элементов [33].*

**Определение 1.11** *Неточное свидетельство — предположение о том, что над подыдеалом  $C'$  задан непротиворечивый фрагмент знаний с интервальными оценками, который определяет вероятности истинности элементов соответствующего подыдеала [33].*

Стоит отметить, что ключевым случаем является пропаганда детерминированного свидетельства, так как на основании предложенного в данном случае уравнения строится алгоритм пропаганды стохастического свидетельства, а также задача линейного программирования для случая неточного и стохастического свидетельств, поэтому в данной работе внимание будет сфокусировано лишь на детерминированном свидетельстве.

В АБС апостериорный вывод подразделяется на две основные задачи. Рассмотрим каждую из них подробнее.

**Определение 1.12** *Первая задача апостериорного вывода заключается в том, чтобы оценить вероятность поступления рассматриваемого свидетельства при уже заданных оценках вероятности истинности элементов фрагмента знаний.*

**Определение 1.13** *Вторая задача апостериорного вывода заключается в том, чтобы вычислить условные вероятности истинности элементов фрагмента знаний при предположении, что имеет место быть свидетельство. Далее, положим, что в рассматриваемый фрагмент знаний поступило свидетельство  $\langle c_i, c_j \rangle$ . Ниже приведем уравнения для решения первой и второй задач апостериорного вывода для фрагментов знаний, построенных над идеалом конъюнктов, идеалом дизъюнктов и множеством пропозиций-квантов.*

Сперва рассмотрим фрагмент знаний над пропозициями-квантами [29]. Решение первой задачи апостериорного вывода запишется следующим образом:

$$p\langle c_i, c_j \rangle = (\mathbf{s}^{\langle i, j \rangle}, \mathbf{P}_q)$$



$$\text{где } \mathbf{s}^{\langle i,j \rangle} = \tilde{\mathbf{s}}_{n-1}^{\langle i,j \rangle} \otimes \tilde{\mathbf{s}}_{n-2}^{\langle i,j \rangle} \otimes \dots \otimes \tilde{\mathbf{s}}_0^{\langle i,j \rangle} \quad \tilde{\mathbf{s}}_k^{\langle i,j \rangle} = \begin{cases} \mathbf{s}^+, & \text{если } x_k \text{ входит в } c_i \\ \mathbf{s}^-, & \text{если } x_k \text{ входит в } c_j \\ \mathbf{s}^0, & \text{иначе} \end{cases}$$

$$\mathbf{s}^+ = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad \mathbf{s}^- = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad \mathbf{s}^0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Уравнение для вектора вероятностей квантов апостериорного вывода выглядит следующим образом [35]:

$$\mathbf{P}_q^{\langle i,j \rangle} = \frac{1}{(\mathbf{s}^{\langle i,j \rangle}, \mathbf{P}_q)} \cdot \mathbf{H}^{\langle i,j \rangle} \mathbf{P}_q.$$

$$\text{где } \mathbf{H}^{\langle i,j \rangle} = \tilde{\mathbf{H}}_{n-1}^{\langle i,j \rangle} \otimes \tilde{\mathbf{H}}_{n-2}^{\langle i,j \rangle} \otimes \dots \otimes \tilde{\mathbf{H}}_0^{\langle i,j \rangle} \quad \tilde{\mathbf{H}}_k^{\langle i,j \rangle} = \begin{cases} \mathbf{H}^+, & \text{если } x_k \text{ входит в } c_i \\ \mathbf{H}^-, & \text{если } x_k \text{ входит в } c_j \\ \mathbf{H}^0, & \text{иначе} \end{cases},$$

$$\mathbf{H}^+ = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{H}^- = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{H}^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Для перехода к условным вероятностям конъюнктов, необходимо умножить полученный вектор  $\mathbf{P}_q^{\langle i,j \rangle}$  на матрицу-перехода  $\mathbf{J}_n$ . Получим решения первой задачи апостериорного вывода для ФЗ над идеалом конъюнктов [33]:

$$p\langle c_i, c_j \rangle = (\mathbf{r}^{\langle i,j \rangle}, \mathbf{P}_c),$$

$$\text{где } \mathbf{r}^{\langle i,j \rangle} = \tilde{\mathbf{r}}_{n-1}^{\langle i,j \rangle} \otimes \tilde{\mathbf{r}}_{n-2}^{\langle i,j \rangle} \otimes \dots \otimes \tilde{\mathbf{r}}_0^{\langle i,j \rangle} \quad \tilde{\mathbf{r}}_k^{\langle i,j \rangle} = \begin{cases} \mathbf{r}^+, & \text{если } x_k \text{ входит в } c_i \\ \mathbf{r}^-, & \text{если } x_k \text{ входит в } c_j \\ \mathbf{r}^0, & \text{иначе} \end{cases},$$

$$\mathbf{r}^+ = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad \mathbf{r}^- = \begin{pmatrix} 1 \\ -1 \end{pmatrix}; \quad \mathbf{r}^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

В принятых обозначениях решением второй задачи апостериорного вывода будет [? ]:

$$\mathbf{P}_c^{\langle i,j \rangle} = \mathbf{J}_n \frac{1}{(\mathbf{s}^{\langle i,j \rangle}, \mathbf{P}_q)} \cdot \mathbf{H}^{\langle i,j \rangle} \mathbf{I}_n \mathbf{P}_q.$$

Определим матрицу  $\mathbf{T}^{\langle i,j \rangle} = \mathbf{J}_n \mathbf{H}^{\langle i,j \rangle} \mathbf{I}_n$ . Тогда полученное выше соотно-

шение для решения второй задачи апостериорного вывода можно записать компактнее:

$$\mathbf{P}_c^{\langle i,j \rangle} = \frac{1}{(\mathbf{r}^{\langle i,j \rangle}, \mathbf{P}_c)} \cdot \mathbf{T}^{\langle i,j \rangle} \mathbf{P}_c$$

где матрица  $\mathbf{T}^{\langle i,j \rangle}$  вычисляется по следующей формуле:

$$\mathbf{T}^{\langle i,j \rangle} = \tilde{\mathbf{T}}_{n-1}^{\langle i,j \rangle} \otimes \tilde{\mathbf{T}}_{n-2}^{\langle i,j \rangle} \otimes \dots \otimes \tilde{\mathbf{T}}_0^{\langle i,j \rangle} \quad \tilde{\mathbf{T}}_k^{\langle i,j \rangle} = \begin{cases} \mathbf{T}^+, & \text{если } x_k \text{ входит в } c_i \\ \mathbf{T}^-, & \text{если } x_k \text{ входит в } c_j \\ \mathbf{T}^0, & \text{иначе} \end{cases},$$

$$\mathbf{T}^+ = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}; \quad \mathbf{T}^- = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}; \quad \mathbf{T}^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Наконец рассмотрим ФЗ над идеалом дизъюнктов. Решением первой задачи апостериорного вывода в таком случае будет следующее уравнение:

$$p\langle c_i, c_j \rangle = (\mathbf{d}^{\langle i,j \rangle}, \mathbf{P}_d').$$

$$\text{где } \mathbf{d}^{\langle i,j \rangle} = \tilde{\mathbf{d}}_{n-1}^{\langle i,j \rangle} \otimes \tilde{\mathbf{d}}_{n-2}^{\langle i,j \rangle} \otimes \dots \otimes \tilde{\mathbf{d}}_0^{\langle i,j \rangle} \quad \tilde{\mathbf{d}}_k^{\langle i,j \rangle} = \begin{cases} \mathbf{d}^+ & , \text{если } x_k \text{ входит в } c_i \\ \mathbf{d}^- & , \text{если } x_k \text{ входит в } c_j \\ \mathbf{d}^0 & , \text{иначе} \end{cases}$$

$$\mathbf{d}^+ = \begin{pmatrix} 1 \\ -1 \end{pmatrix}; \quad \mathbf{d}^- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad \mathbf{d}^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Аналогично случаю с идеалом конъюнктов и множеством пропозиций-квантов решение второй задачи апостериорного вывода записано ниже:

$$\mathbf{P}_d'^{\langle i,j \rangle} = \frac{1}{(\mathbf{d}^{\langle i,j \rangle}, \mathbf{P}_d')} \mathbf{M}^{\langle i,j \rangle} \times \mathbf{P}_d'$$

$$\text{где } \mathbf{M}^{\langle i,j \rangle} = \widetilde{\mathbf{M}}_{n-1}^{\langle i,j \rangle} \otimes \widetilde{\mathbf{M}}_{n-2}^{\langle i,j \rangle} \otimes \dots \otimes \widetilde{\mathbf{M}}_0^{\langle i,j \rangle} \quad \widetilde{\mathbf{M}}_k^{\langle i,j \rangle} = \begin{cases} \mathbf{M}^+, & \text{если } x_k \text{ входит в } c_i \\ \mathbf{M}^-, & \text{если } x_k \text{ входит в } c_j \\ \mathbf{M}^0, & \text{иначе} \end{cases}$$

$$\mathbf{M}^+ = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}; \quad \mathbf{M}^- = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}; \quad \mathbf{M}^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

## 1.4. Выводы по главе

В данной главе были введены ряд определений, объектов и фактов из теории АБС, а также рассмотрены некоторые аспекты, связанные с вероятностной семантикой логико-вероятностных моделей фрагментов знаний над идеалами конъюнктов, дизъюнктов и пропозициями-квантами. Описаны ключевые формулы, используемые в локальном апостериорном выводе для каждого из трех случаев, а также при проверке непротиворечивости данных, при этом показаны переходы между от одного типа представления фрагмента знаний к другому при решении задачи апостериорного вывода.

## 2. Алгебраические байесовские сети: база данных

### 2.1. Введение

В настоящее время активно развивается теоретический аппарат алгебраических байесовских сетей. Получены такие результаты по теории АБС, как: исследованы АБС, в которых фрагменты знаний могут быть построены не только над идеалом конъюнктов, но и над идеалом дизъюнктов и множеством пропозиций-квантов [29, 32, 33, 34, 35]. Также разрабатываются и реализуются программные решения для работы с АБС [19, 30]. Но все они не содержат аппарата для долгого хранения данных в системе — базы данных или не учитывают новейшие теоретические достижения в силу своего возраста [33].

В данной главе представлена схема разработанной базы данных для хранения таких объектов, как свидетельства, фрагменты знаний и их составляющих, самих алгебраических байесовских сетей. Также приведено обоснование выбора технологии, средствами которой велась разработка (MS SQL Server [20]). Кроме того описаны проблемы и недостатки существующих решений, а также их решения, которые были предложены и реализованы в ходе разработки.

Разработанная база данных является частью проекта «AlgBN Web App» по вебизации АБС, который представляет собой web-платформу для обучения теории алгебраических байесовских сетей.

### 2.2. Выбор технологий

При реализации базы данных была использована реляционная система управления базами данных (СУБД) Microsoft SQL Server [20]. По ряду причин, приведенных ниже, была выбрана именно реляционная система управления базами данных (РСУБД), а не NoSQL. Ниже приведены причины выбора SQL базы данных, основанные на [12, 16, 23].

Во-первых, NoSQL базы данных работают с не регламентированны-

ми структурами данных (слабо типизированными). Разрабатываемая база данных имеет сложную, но строгую структуру, так как хранит в себе такие объекты как АБС, ФЗ и свидетельства, структура которых строго определена теоретическим аппаратом АБС, которые удобно описывать через систему отношений (реляционная модель). Из-за слабой структурированности данных в NoSQL базах данных у них существует проблема накладных расходов в коде приложения при смене модели данных, что несомненно важно учитывать в нашем случае, так как развитие теоретического аппарата ведет к постоянной неизбежной смене модели данных. Также слабо типизированные данные означают отсутствие всевозможных ограничений со стороны базы (таких как `not null`, `unique`, `check constraint` и т.д.), которые также необходимы для хранения информации в создаваемой базе в необходимом виде на уровне самой базы, что возможно на должном уровне лишь в реляционной модели.

Во-вторых, NoSQL базы данных представляют хранимые данные в виде агрегатов, то есть оперируют с сущностями как с целостными объектами, в отличие от SQL решений, которые сохраняют логическую бизнес-сущность приложения в различные физические таблицы в целях нормализации. Таким образом, существует возможность получения частичных объектов из базы данных при использовании SQL-модели (например, получение конкретного конъюкта из фрагмента знаний у определенной АБС, а не весь ФЗ целиком). Нормализация данных имеет следующие преимущества перед агрегатами (а, следовательно, и использование реляционной модели данных):

- При обновлении информации сохраняется целостность данных, то есть необходимо изменить запись лишь в одной таблице, а не в нескольких.
- Нормализованные данные ориентированы на широкий спектр запросов к данным.
- Данные в виде агрегатов (используемые в NoSQL базах данных) могут быть оптимизированы только под определенные виды за-

просов, а также не приспособлены к простому и быстрому обновлению.

Стоит отметить, что NoSQL базы данных наиболее эффективны в распределенной среде, но так как внедрение базы данных предполагает внедрение в проект без распределенной нагрузки, то данное преимущество не рассматривается в рамках данной работы.

В-третьих, разрабатываемое приложение для работы с АБС, в рамках которого ведется разработка базы данных, рассчитано на продолжительную работу пользователей с возможностью сохранения промежуточного результата в системе. При этом необходимо гарантировать выполнение транзакций для избежания внутренних ошибок и конфликтов. NoSQL технология является достаточно новой, поэтому работа с транзакциями в них не стабильна. Другими словами, NoSQL базы данных имеют слабые ACID свойства и поэтому менее привлекательны в данном проекте.

В-четвертых, выбранная технология поддерживает высокую скорость работы на запись и обновление данных, что, с учетом работы с большими объемами табличных данных, несомненно важно в разрабатываемой системе. Помимо этого реляционные базы данных являются довольно хорошо изученной технологией и в настоящее время имеются качественно задокументированные как платные, так и бесплатные решения с поддержкой.

Подводя итог вышесказанного, в рамках данной работы была выбрана реляционная система управления базами данных. В данный момент популярны несколько РСУБД, такие как MySQL [21], SQL Server [20] и PostgreSQL [24]. Рассмотрим каждую из них.

MySQL [21] в настоящее время широко используется, например на таких популярных ресурсах, как Wikipedia, YouTube, Facebook. Также MySQL имеет открытый исходный код и при распространении с открытым исходным кодом ее использование является бесплатным. В настоящее время MySQL адаптирована под большое количество операционных систем, таких как Windows, Unix-подобные системы, OS/2. Последняя выпущенная версия MySQL была в августе в 2012 году, что

означает, что в настоящее время MySQL не имеет активной поддержки со стороны разработчиков. С точки зрения клиентского программного обеспечения MySQL не уступает современным СУБД, так как имеет как утилиты для работы из командной строки, так и GUI (пользовательский графический интерфейс) для работы с графической оболочкой. Сжатые данные в MySQL доступны только для чтения. При записи происходит распаковка таблицы, затем запись, затем упаковка. При этом даже такая поддержка есть только в движке MyISAM. Кроме того некоторая функциональность у некоторых типов таблиц является взаимно исключающей. Например, если необходимо одновременно иметь возможность полнотекстового поиска и поддержку внешних ключей, то такую функциональность реализовать не получится. Это связано с тем, что полнотекстовый поиск доступен только на движке MyISAM, а внешние ключи доступны только на движке InnoDB, а таблица не может быть сразу этих двух типов. Существует поддержка платформы .NET средствами ADO.Net-драйвера, что безусловно важно. Так как математическая библиотека для работы с АБС и с базой данных написана на языке C#.

PostgreSQL [24] имеет открытый исходный код и при распространении с открытым исходным кодом использование данной СУБД является бесплатным. Кроме того PostgreSQL адаптирована под большое количество операционных систем, таких как Windows, Unix-подобные систем. PostgreSQL в настоящее время имеет активную техническую поддержку. Последняя выпущенная версия приложится на февраль 2017 года, что несомненно является преимуществом. Как и свои главные конкуренты, PostgreSQL имеет утилиты для работы из командной строки, а также графическую оболочку для пользователей. В настоящее время в рассматриваемой СУБД существует возможность постраничного сжатия данных. Для работы с .NET приложениями существует такое средство как .Net Data Provider for Postgresql. Что делает возможным использование PostgreSQL в разрабатываемом проекте.

SQL Server [20] является платной технологией (кроме версии Express Edition и пробной версией Evaluation Edition). Но при этом SQL Server

предоставляет возможность бесплатного использования в рамках студенческой программы DreamSpark с последующим бесплатным размещением на сервере Microsoft Azure (с ограничением объема базы данных). SQL Server распространен только под операционной системой Windows, но это ограничение является приемлемым в рамках реализуемого проекта. Последний релиз SQL Server приходится на ноябрь 2016 года, что было совсем недавно. При этом каждые год-два выходят новые версии, что позволяет говорить об активной поддержке данной СУБД со стороны разработчиков. SQL Server не является исключением и также имеет все необходимые средства для работы пользователя с точки зрения клиентско-программного обеспечения: утилиты для работы с командной строки, а также графическую оболочку. SQL Server, начиная с версии 2008 года, поддерживает прозрачное сжатие данных без каких-либо ограничений. Также помимо обычной SQL-аутентификации, поддерживаемой практически всеми современными реляционными СУБД, существует возможность Windows-аутентификации. Как уже было сказано выше, математическая библиотека реализована на языке C#, поэтому необходима поддержка платформы .NET, которая присутствует в среде SQL Server с помощью платформы ADO.Net.

По функциональности все три СУБД (MySQL, SQL Server и PostgreSQL) являются схожими. Каждая из них лучше в некоторых аспектах, в некоторых хуже. Но в целом функциональность и производительность находится примерно на одном уровне. Было принято решение выбирать между наиболее поддерживаемыми в настоящий момент средствами — SQL Server и PostgreSQL. SQL Server в отличие от PostgreSQL дает возможность бесплатного хостинга в рамках студенческой программы DreamSpark на сервере Microsoft Azure было принято решение выбрать именно эту реляционную СУБД для создания базы данных.

### **2.3. Схема базы данных**

На Рис. 2 представлена разработанная структура базы данных, хранящей информацию об АБС.



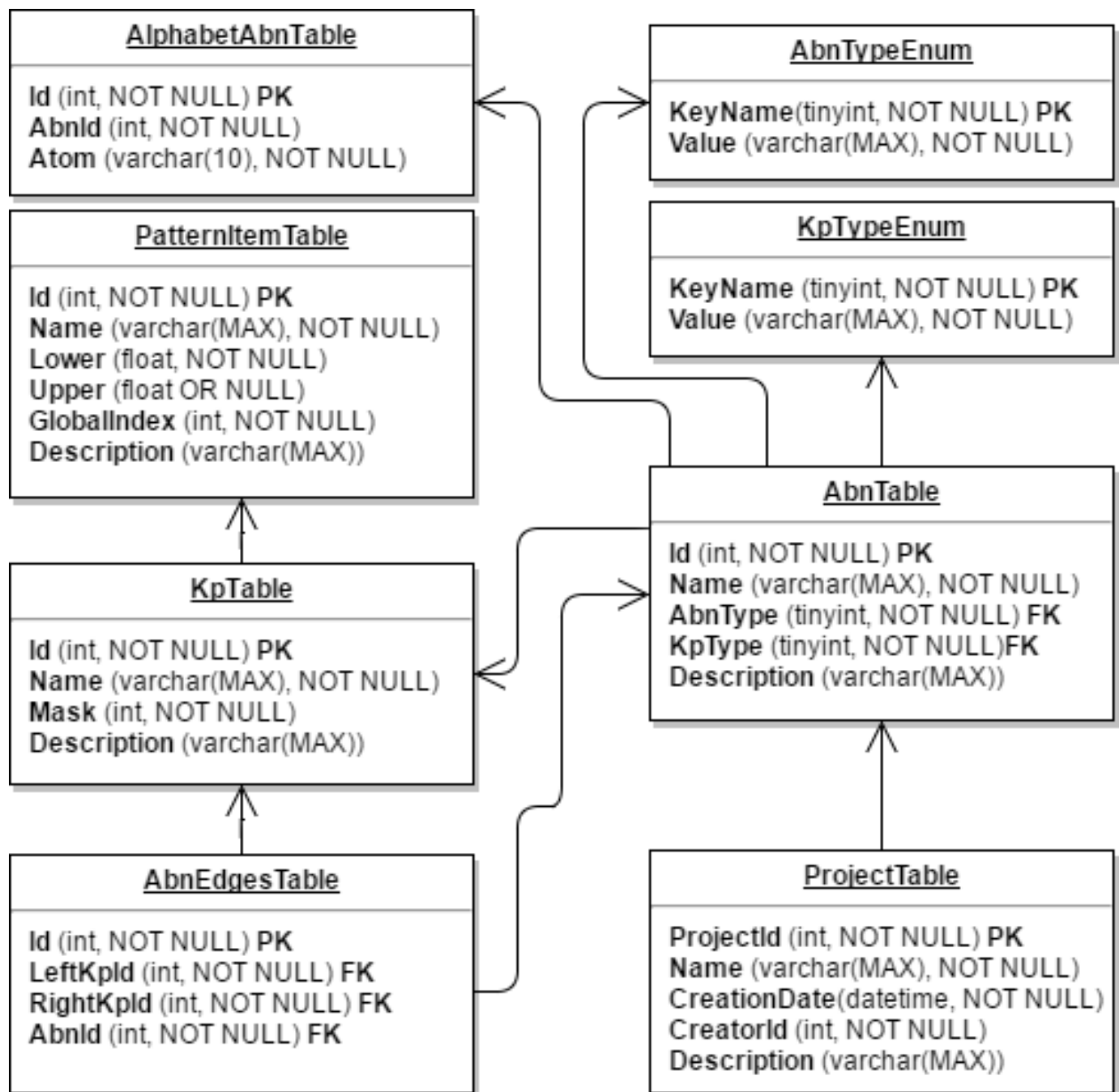


Рис. 2: Представление АБС в базе данных

Все АБС объединяются в проекты, где проекты представлены в таблице ProjectTable. Каждый проект имеет уникальный идентификатор (ProjectId), имя (Name), дату создания (CreationDate) и создателя проекта (CreatorId). Такая система группировки позволит в дальнейшем при реализации пользовательского интерфейса добиться наилучшего представления данных.

Алфавит, над которым построена АБС хранится в таблице AlphabetAbnTable поэлементно. Каждый атом алфавита имеет уникальный идентификатор (Id), уникальный идентификатор АБС (AbnId),

алфавиту которой он принадлежит и имя самого элемента (Atom), которое ограничено 10 символами.

Таблицы AbnTypeEnum и KrTypeEnum имеют одинаковую структуру (как и все таблицы с префиксом Enum). Данные таблицы являются аналогом словаря, в котором номеру типа - целому числу - сопоставляется его текстовый эквивалент. Таблица AbnTypeEnum содержит в себе информацию, над какими множеством построены ФЗ в АБС (идеал дизъюнктов, идеал конъюнктов или множество пропозиций-квантов). Таблица KrTypeEnum характеризует тип векторов вероятностей истинности элементов ФЗ (скалярные или интервальные значения).

АБС в базе данных представлена в таблице AbnTable, которая содержит такие характеристики как уникальный идентификатор (Id), имя (Name), ссылки на таблицы AbnTypeEnum и KrTypeEnum (поля AbnType и KrType соответственно).

ФЗ представлены в таблице KrTable и имеют уникальный идентификатор (Id), имя (Name) и глобальный индекс - маску (Mask), который позволяет определить, над каким набором атомов алфавита АБС построен данный ФЗ.

Элементы ФЗ (конъюнкты, дизъюнкты или кванты) содержатся в таблице PatternItemTable, так как имеют общую структуру, как было сказано выше. Они характеризуются уникальным идентификатором (Id), именем (Name), нижним и верхним значением оценки вероятностей (Lower и Upper соответственно) и глобальным индексом (GlobalIndex), являющимся аналогом поля Mask у ФЗ. Он позволяет определить, над какими атомами алфавита ФЗ построен PatternItem.

АБС можно представить в виде вторичной структуры. В работе[[ представлена программная реализация разработанных алгоритмов для работы со вторичной структурой АБС. Приложение позволяет генерировать множество минимальных графов смежности вторичной структуры. Таблица AbnEdgesTable хранит одну из таких структур по выбору пользователя. Граф смежности в данном случае представляется набором ребер, каждое из которых имеет уникальный идентификатор (Id), ссылки на левую и правую вершину - ФЗ, которые соединяет дан-

ное ребро (LeftKpId и RightKpId соответственно) и уникальный идентификатор АБС, к которой относится данное ребро графа смежности (AbnId).

На Рис. 3 представлена структура свидетельства. Тип свидетельств (детерминированное, стохастическое и неточное) хранится в таблице EvidenceTypeEnum. Таблица EvidenceTable содержит основную информацию о свидетельстве: уникальный идентификатор (Id), уникальный идентификатор АБС, которой оно принадлежит (AbnId), имя (Name), глобальный индекс, позволяющий определить атомы алфавита, над которыми построено свидетельство (GlobalIndex), ссылку на таблицу EvidenceTypeEnum для определения типа свидетельства (Type) и локальный индекс (LocalIndex) для детерминированного свидетельства. Стохастические и неточные свидетельства состоят из элементов, которые содержатся в таблице EvidencePatternItemTable и являются полными аналогами PatternItem-ов.

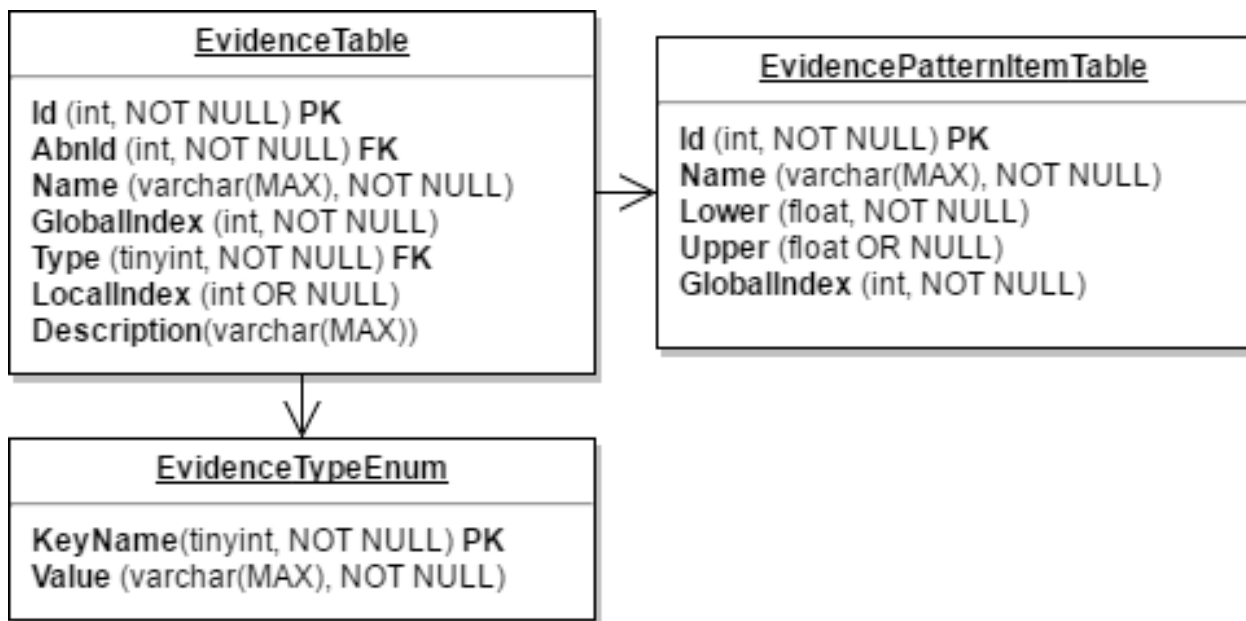


Рис. 3: Представление свидетельства в базе данных

Связи «один ко многим», например «Набор ФЗ у конкретной АБС» реализованы с помощью дополнительных таблиц с префиксом Link, которые содержат по два поля — уникальных идентификаторов объектов, состоящих в таком отношении. Данные таблицы исключены из представленных на Рис. 2, 3 схемах для компактного представления.

Такие объекты, как АБС, ФЗ, элементы ФЗ (PatternItem), свидетельство и проект содержат необязательное поле Description для комментария.

На Рис. 4 представлена полная схема разработанной базы данных, включающая связующие Link таблицы.

## 2.4. Существующие проблемы и их решение

Как было сказано выше, АБС состоит из набора ФЗ, которые могут быть построены как над идеалом конъюнктов, так и над идеалом дизъюнктов или множеством пропозиций-квантов. Оценки вероятности истинности элементов ФЗ могут быть скалярными или интервальными, но во всех случаях общая структура АБС и ее элементов остается неизменной, меняются только алгоритмы обработки данных. Для решения проблемы унификации данных различных типов тип ФЗ был вынесен в таблицы AbnTypeEnum и KpTypeEnum, которые хранят информацию о типе ФЗ на уровне АБС. При этом не происходит дублирования информации, так как теорией подразумевается, что в рамках одной АБС не может быть ФЗ разных типов.

Для сохранения единства структуры у ФЗ с скалярными и интервальными оценками вероятности истинности элементов было принято решение не разделять данные случаи в несколько таблиц, так как скалярные оценки являются частным случаем интервальных с одинаковыми значениями вероятности нижней и верхней границ. Для избежания дублирования оценок в случае скалярных оценок верхняя оценка хранит пустое значение (NULL).

Так как АБС состоит из нескольких связанных между собой ФЗ, где пересечение ФЗ(сепаратор) является также ФЗ, построенным над элементами, принадлежащими обоим ФЗ, то возникла проблема дублирования элементов, принадлежащих сепаратору. Для решения данной проблемы было принято решение хранить не сами ФЗ, а элементы, из которых они состоят (конъюнкты, дизъюнкты, кванты). Таким образом, ФЗ содержит ссылки на нужные элементы, из которых он состоит,

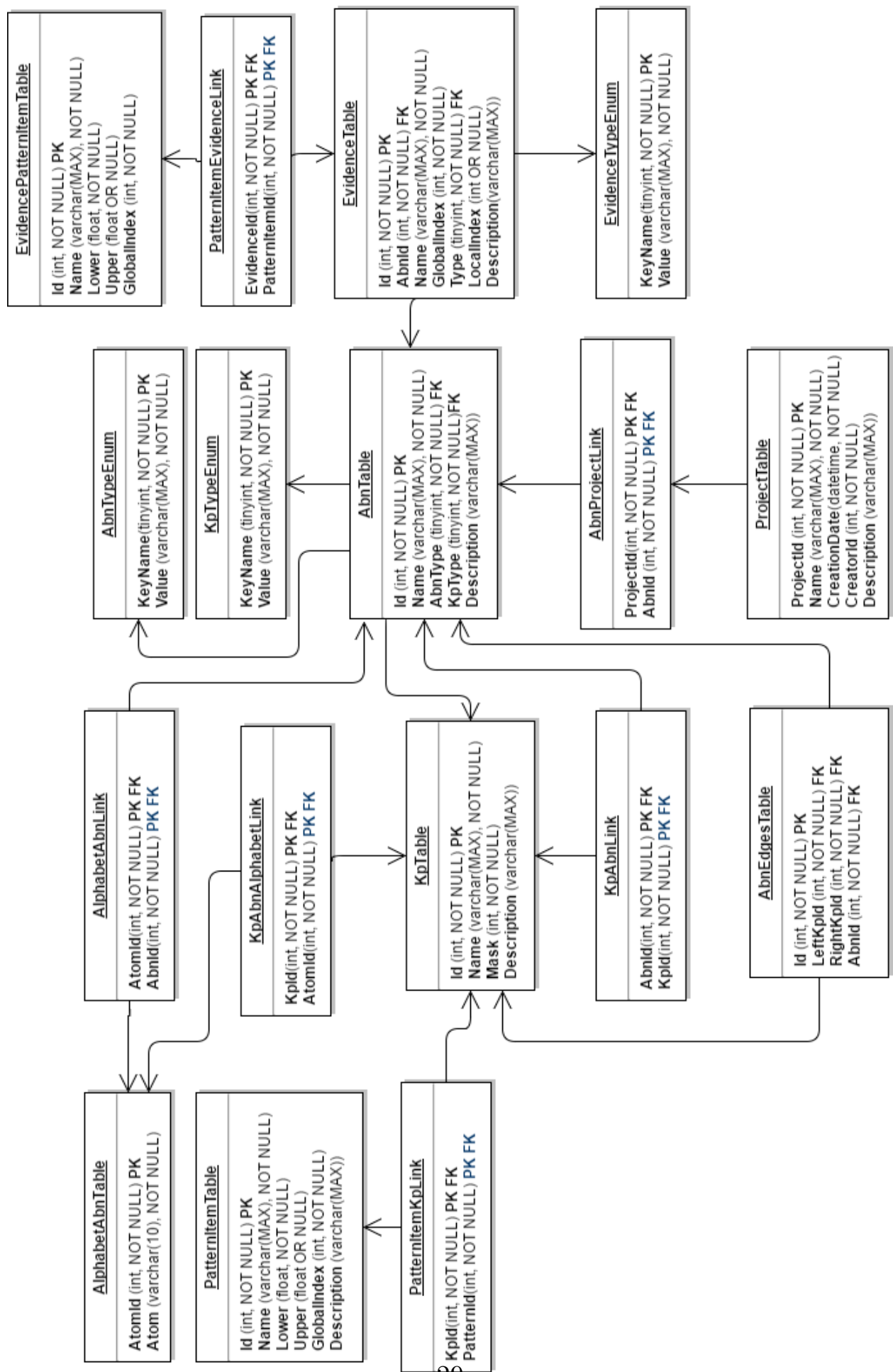


Рис. 4: Схема разработанной базы данных

а на элемент из пересечения ФЗ ссылаются несколько ФЗ.

При хранении алфавита, над которым построена АБС, необходимо учитывать порядок элементов для корректной работы алгоритмов. Было решено не вводить дополнительного поля с порядковым номером элемента алфавита, так как у каждого символа есть уникальный идентификатор (id), который позволит сохранить нужную последовательность элементов алфавита, используя сортировку по возрастанию id нужных элементов.

В задачах апостериорного вывода в АБС используются три вида свидетельств: детерминированное, стохастическое и неточное. Детерминированное свидетельство состоит из элементов, вероятности которых могут быть 0 или 1, поэтому данный тип свидетельства удобно представлять в виде двоичной маски, нежели хранить все элементы свидетельства целиком, занимая тем самым значительно больше места. В целях улучшения производительности системы и минимизации требуемого объема памяти для хранения таких свидетельств в таблицу EvidenceTable было добавлено поле LocalIndex, которое хранит значение — двоичную маску или пустое значение (NULL) в случае стохастического и неточного свидетельств.

## 2.5. Выводы по главе

В данной главе был проведен сравнительный анализ SQL и NoSQL технологий и аргументирован выбор реляционной СУБД MS SQL Server. Также была представлена и подробно описана разработанная база данных с описанием полей всех таблиц. Кроме того, были описаны возникшие проблемы при разработке и реализации и методы их решения

## 3. Расширенная документация:

### материалы для методического пособия

#### 3.1. Введение

В настоящее время для работы с АБС существует комплекс программ «AlgBN Math Library» [19, 30] на языке C#. Он включает в себя средства для решения задач локального логико-вероятностного вывода (ЛВВ), а также средства для проверки и поддержания непротиворечивости. Данный комплекс программ активно используется в web-приложении «AlgBN Web App» [28], так как в нем заложен основной теоретический аппарат по АБС. Но в силу объемности комплекса, его изучение для последующего реинжиниринга становится трудоемкой задачей. В связи с этим было принято решение создания методического пособия для работы с данным комплексом программ, которое также содержало бы основные теоретические выкладки по каждому разделу. В связи с этим были подготовлены все необходимые материалы для такого рода пособия.

При создании материалов для методического пособия, разработанные примеры проверялись двумя способами: аналитически и программно. Аналитическая проверка велась частично вручную, частично с использованием среды MAPLE (для автоматизации решения задач линейного программирования). Программная проверка велась с помощью комплекса программ на языке C#.

#### 3.2. Структура

В рамках данной работы были разработаны материалы к методическому пособию к комплексу программ «AlgBN Math Library», которые включают в себя следующие виды логико-вероятностного вывода в АБС:

- Проверка и поддержание непротиворечивости
  - Проверка непротиворечивости в скалярном ФЗ

- Поддержание непротиворечивости в интервальном ФЗ
- Априорный вывод
  - Задача априорного вывода в скалярном ФЗ
  - Задача априорного вывода в интервальном ФЗ
- Апостериорный вывод
  - Пропагирование детерминированного свидетельства в ФЗ со скалярными оценками
  - Пропагирование детерминированного свидетельства в ФЗ с интервальными оценками
  - Пропагирование стохастического свидетельства в ФЗ со скалярными оценками
  - Пропагирование стохастического свидетельства в ФЗ с интервальными оценками
  - Пропагирование неточного свидетельства в ФЗ со скалярными оценками
  - Пропагирование неточного свидетельства в ФЗ с интервальными оценками

Каждый вид логико-вероятностного вывода (ЛВВ) включает в себя следующее:

- Краткая теоретическая сводка с необходимыми формулами по данному виду ЛВВ
- Постановка конкретной задачи (с графической иллюстрацией исходных данных)
- Аналитическая проверка (разработанные численные эксперименты по данному виду ЛВВ)
- Программная проверка (разработанные по проведенным экспериментам тесты в комплексе программ)



- Сравнение аналитических и программных результатов (программно)

### 3.3. Аналитическая проверка: численные эксперименты

Для проведения аналитической проверки были разработаны и проведены численные эксперименты по разным видам ЛВВ в АБС. Каждый из разработанных примеров сопровождается подробными математическими выкладками, которые позволяют поэтапно проследить получение результатов.

При аналитической проверке все вычисления велись вручную и средствами среды MAPLE. В частности все задачи линейного программирования (ЗЛП) были решены в среде MAPLE в целях автоматизации эксперимента. Стоит отметить, что разработка экспериментов велась над алфавитом  $\{X_1, X_2\}$ .

Подробно ознакомиться со всеми описанными экспериментами можно в приложении, которые прилагаются к данной дипломной работе.

Приведем пример аналитической проверки.

Возьмем фрагмент знаний над алфавитом  $A_c\{x_1, x_2\}$  :

$$P_c = \begin{pmatrix} 1 \\ p(x_1) \\ p(x_2) \\ p(x_2x_1) \end{pmatrix} = \begin{pmatrix} 1 \\ [0.5, 0.9] \\ [0.8, 0.87] \\ [0.5, 0.74] \end{pmatrix}$$

Его можно также увидеть на Рис. 5. Он является согласованным.

Будем пропагировать детерминированное свидетельство:  $\langle x_2, x_1 \rangle = \langle 10, 01 \rangle$

$$T^{\langle i, j \rangle} = \widetilde{T_{n-1}^{\langle i, j \rangle}} \otimes \widetilde{T_{n-2}^{\langle i, j \rangle}} \otimes \dots \otimes \widetilde{T_0^{\langle i, j \rangle}}$$

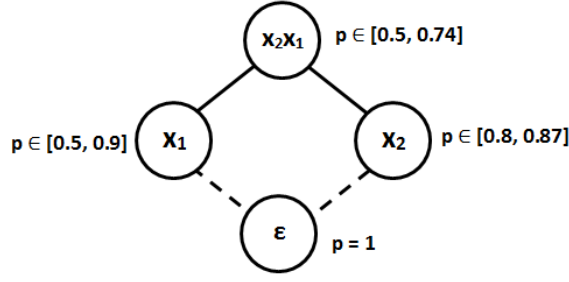


Рис. 5: Фрагмент знаний со скалярными оценками

$$\widetilde{T_k^{(i,j)}} = \begin{cases} T^+ & \text{если } x_k \text{ входит в } c_i, \\ T^- & \text{если } x_k \text{ входит в } c_j, \\ T^0 & \text{иначе.} \end{cases}$$

$$T^+ = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$T^- = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$$

$$T^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

**Первая задача апостериорного вывода.**

Будем решать соответствующие ЗЛП:

$$p(\langle i, j \rangle) = \min_{D \cup \varepsilon} \setminus \max_{D \cup \varepsilon} (T^{(i,j)} \times P_c)[0] \quad (1)$$

$$T^{\langle 10, 01 \rangle} = T^+ \otimes T^- = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$(T^{\langle 10, 01 \rangle} \times P_c)[0] = p(x_2) - p(x_2x_1)$$

$$p(\langle 10, 01 \rangle)^- = 0.06$$

$$p(\langle 10, 01 \rangle)^+ = 0.37$$

Таким образом,

$$p(\langle 10, 01 \rangle) \in [0.06, 0.37]$$

**Вторая задача апостериорного вывода.**

Будем решать серию ЗЛП:

$$\min \setminus \max T^{\langle i, j \rangle} \times D$$

при условиях:

$$\{\lambda P_c^- \leq D\} \cup \{D \leq \lambda P_c^+\} \cup \{I_2 \times D \geq 0\} \cup \{(L^{\langle i, j \rangle}, D) = 1\} \cup \{\lambda \geq 0\}$$

Введем замену переменных:

$$D = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}$$

Получим следующие ограничения для ЗЛП:

$$a_1 - a_2 - a_3 + a_4 \geq 0$$

$$a_2 - a_4 \geq 0$$

$$a_3 - a_4 \geq 0$$

$$a_4 \geq 0$$

$$a_3 - a_4 = 1$$

$$a_1 - \lambda = 0$$

$$a_2 - 0.5 \cdot \lambda \geq 0$$

$$a_3 - 0.8 \cdot \lambda \geq 0$$

$$a_4 - 0.5 \cdot \lambda \geq 0$$

$$a_2 - 0.9 \cdot \lambda \leq 0$$

$$a_3 - 0.87 \cdot \lambda \leq 0$$

$$a_4 - 0.74 \cdot \lambda \leq 0$$

Целевая функция будет иметь вид::

$$a_3 - a_4 \longrightarrow \min \setminus \max$$

После решения серии ЗЛП, получим решением второй задачи апосте-

риорного вывода:

$$P_c^{\langle 10, 01 \rangle} = \begin{pmatrix} 1 \\ [0.3, 0.6] \\ [0.575, 0.65] \\ [0.225, 0.445] \end{pmatrix}$$

### 3.4. Программная проверка

На основе проведенных численных экспериментов были написаны тесты к комплексу программ «AlgBN Math Library» по выше указанным видам локального ЛВВ. Для этого в каждой из машин (Inferrer и Propagator) были созданы классы для каждого вида проверок:

- Проверка и поддержание непротиворечивости:
  - Test\_Consistent
  - Test\_Consistent\_Interval
- Априорный вывод:
  - Test\_Priori\_Conclusion\_Scalar
  - Test\_Priori\_Conclusion\_Interval
- Апостериорный вывод:
  - Test\_Propagate\_Det\_Scalar
  - Test\_Propagate\_Det\_Interval
  - Test\_Propagate\_Stoch\_Scalar
  - Test\_Propagate\_Stoch\_Interval
  - Test\_Propagate\_Imprecise\_Scalar
  - Test\_Propagate\_Imprecise\_Interval

В методических материалах приведен текст методов из данных классов с подробными комментариями, что позволяет ознакомиться с функционалом комплекса программ напрямую. После аналитической и про-

граммной проверок была организована проверка совпадения результатов в каждом из классов. При данной проверке задавались ожидаемые результаты (результаты, полученные при аналитической проверке) и запускался комплекс программ на разработанных тестах. Стоит отметить, что все тесты прошли успешную проверку, подтверждение которой можно увидеть в методическом пособии в виде скриншотов результата выполнения программы.

Покажем пример программной проверки для фрагмента знаний, приведенного ранее и изображенного на Рис. 5.

### **Программная проверка 1 и 2 задач апостериорного вывода.**

```

1  Метод для проверки 1 и 2 задач апостериорного вывода
2  public static bool Test_Propagate_Det_Interval_1()
3  {
4      Переменные флаги для проверки результатов 1 и 2 задач
5      var test_1task = false; var test_2task = true;
6
7      Ожидаемый результат выполнения 1 задачи нижняя( и верхняя границы)
8      var answer_1taskL = 0.06; var answer_1taskU = 0.37;
9
10     Ожидаемый результат выполнения 2 задачи нижняя( и верхняя границы)
11     double[] answer_2taskL = {1, 0, 1, 0};
12     double[] answer_2taskU = { 1, 0, 1, 0 };
13
14     Создаем интервальный ФЗ с глобальным индексом 11 детерминированное(
свидетельство)
15     IntervalConjKP_I kp = new IntervalConjKP(Convert.ToInt64("11", 2),
new [] {1, 0.5, 0.8, 0.5}, new [] {1, 0.9, 0.87, 0.74});
16
17     Создаем машину пропагатор и добавляем ФЗ
18     DeterministicIntervalConjunctsLocalPropagatorI scprop = new
DeterministicIntervalConjunctsLocalPropagator(kp);
19
20     Создаем бинарный ФЗ с глобальным индексом 11
21     BinaryConjKP evid = new BinaryConjKP(Convert.ToInt64("11", 2));
22
23     Добавляем детерминированное свидетельство
24     evid.SetQuant(Convert.ToInt32("10", 2));
25
26     Проверяем полученные результаты 2 задачи
27     if (scprop.propagate(evid))
28     {
29         test_1task = CompareEq(answer_1taskL, scprop.

```

```

30         getEvidenceProbabilityLB()) &&
           CompareEq(answer_1taskU, scprop.
31         getEvidenceProbabilityUB());
32     }
33     Проверяем полученные результаты 2 задачи
34     var resL = scprop.getResult().GetBoundsL();
35     var resU = scprop.getResult().GetBoundsU();
36     for (var i = 0; i < resL.Length; i++)
37         if (!CompareEq(resL[i], answer_2taskL[i]))
38             test_2task = false;
39
40     for (var i = 0; i < resU.Length; i++)
41         if (!CompareEq(resU[i], answer_2taskU[i]))
42             test_2task = false;
43
44     Возвращаем результаты проверок обеих задач
45     return test_1task && test_2task;
46 }
47
48 Метод для обработки результатов
49 public static void PrintTest()
50 {
51     Шаблон имени методов тестов—
52     string name = "Test_Propagate_Det_Interval_";
53
54     Общее количество тестов
55     int countTest = 1;
56
57     Далее в цикле вызов каждого из методов тестов—
58     for (int i = 1; i <= countTest; i++)
59     {
60         string curName = name + i;
61         Test_Propagate_Det_Interval obj = new
Test_Propagate_Det_Interval();
62         MethodInfo method = obj.GetType().GetMethod(curName);
63         bool result = (bool)method.Invoke(null, null);
64
65         Если тест успешно завершен, то говорим об успехе
66         Иначе говорим о провале
67         if (result)
68             Console.WriteLine(curName + " is success!");
69         else
70             Console.WriteLine(curName + " is unsuccessful!");
71     }

```

### Листинг 1: Программная проверка

Также в конце каждого типа проверки происходит сравнение результатов аналитической и программной проверок. После запуска метода с листинга 1, можно увидеть результаты на консоли об успешном прохождении эксперимента.

## 3.5. Выводы по главе

В данной главе была описана необходимость создания методического пособия к комплексу программ «AlgBN Math Library». Кроме того подробно описана структура разработанных материалов, используемые технологии, а также показаны примеры из них.

## Заключение

В настоящей работе были представлены базовые определения, связанные с алгебраическими байесовскими сетями, приведены ссылки на аналоги разработанных схем баз данных с указанными недостатками, описан выбор соответствующих технологий для реализации базы данных. Кроме того была приведена подробная схема разработанной базы данных и проанализированы и описаны все возникшие проблемы при разработке, а также указано их решение. Также приведены разработанные материалы, сопуствующие созданию методического пособия для работы с комплексом программ «AlgBN Math Library».

В результате проделанной работы была разработана и реализована база данных для хранения АБС и работы с ними, что позволило интегрировать ее в web-платформу «AlgBN Web App» по обучению теории АБС. Также разработанная база данных учитывает все недостатки прошлых аналогов и новейшие теоретические результаты в данной области. Кроме того были разработаны материалы для методического пособия к комплексу программ «AlgBN Math Library», которые содержат не только подробное руководство использования данного комплекса, а также краткие теоретические аспекты для лучшего понимания темы. Данные материалы позволят выпустить методическое пособие, которое необходимо при знакомстве с данной тематикой.



## Список литературы

- [1] S. Arangio and F. Bontempi. Structural health monitoring of a cable-stayed bridge with bayesian neural networks. *Structure and Infrastructure Engineering*, 11(4):575–587, 2015.
- [2] O. Arsene, I. Dumitrache, and I. Mihiu. Expert system for medicine diagnosis using software agents. *Expert Systems with Applications*, 42(4):1825–1834, 2015.
- [3] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.
- [4] J. Borras, A. Moreno, and A. Valls. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, 41(16):7370–7389, 2014.
- [5] Z. Cai, P. Guo, S. Si, Z. Geng, C. Chen, and L. Cong. Analysis of prognostic factors for survival after surgery for gallbladder cancer based on a bayesian network. *Scientific Reports*, 7(1):293, 2017.
- [6] A. Constantinou, N. Fenton, and M. Neil. pi-football: A bayesian network model for forecasting association football match outcomes. *Knowledge-Based Systems*, 36:322–339, 2012.
- [7] J. Dabrowski and J. de Villiers. Maritime piracy situation modelling with dynamic bayesian networks. *Information fusion*, 23:116–130, 2015.
- [8] P. Dawson, R. Gailis, and A. Meehan. Detecting disease outbreaks using a combined bayesian network and particle filter approach. *Journal of theoretical biology*, 370:171–183, 2015.
- [9] X. Desforges, M. Diévar, and B. Archimède. A prognostic function for complex systems to support production and maintenance co-operative planning based on an extension of object oriented bayesian networks. *Computers in Industry*, 86:34–51, 2017.

- [10] S. Dragicevic, S. Celar, and M. Turic. Bayesian network model for task effort estimation in agile software development. *Journal of Systems and Software*, 127:109–119, 2017.
- [11] S. Hamilton, C. Pollino, and A. Jakeman. Habitat suitability modelling of rare species using bayesian networks: Model evaluation under limited data. *Ecological Modelling*, 299:64–78, 2015.
- [12] J. Han, E. Haihong, G. Le, and J. Du. Survey on nosql database. In *Pervasive computing and applications (ICPCA), 2011 6th international conference on*, pages 363–366. IEEE, 2011.
- [13] J-S. Kim and C-H. Jun. Ranking evaluation of institutions based on a bayesian network having a latent variable. *Knowledge-Based Systems*, 50:87–99, 2013.
- [14] YG. Kim, SM. Lee, and PH. Seong. A methodology for a quantitative assessment of safety culture in npps based on bayesian networks. *Annals of Nuclear Energy*, 102:23–36, 2017.
- [15] S-S. Leu and C-M. Chang. Bayesian-network-based fall risk evaluation of steel construction projects by fault tree transformation. *Journal of Civil Engineering and Management*, 21(3):334–342, 2015.
- [16] Y. Li and S. Manoharan. A performance comparison of sql and nosql databases. In *Communications, Computers and Signal Processing (PACRIM), 2013 IEEE Pacific Rim Conference on*, pages 15–19. IEEE, 2013.
- [17] KF-R. Liu, J-Y. Kuo, Ken. Yeh, C-W. Chen, H-H. Liang, and Y-H. Sun. Using fuzzy logic to generate conditional probabilities in bayesian belief networks: a case study of ecological assessment. *International Journal of Environmental Science and Technology*, 12(3):871–884, 2015.
- [18] Q. Mahboob, E. Schöne, U. Maschek, and J. Trinckauf. Investment into human risks in railways and decision optimization. *Human and*

*Ecological Risk Assessment: An International Journal*, 21(5):1299–1313, 2015.

- [19] E. Mal'chevskaya, A. Berezin, A. Zolotin, and A. Tulupyev. Algebraic bayesian networks: Local probabilistic-logic inference machine architecture and set of minimal joint graphs. In *Proceedings of the First International Scientific Conference "Intelligent Information Technologies for Industry"(IITI'16)*, pages 69–79. Springer, 2016.
- [20] Microsoft. Sql server technical documentation. URL: <https://docs.microsoft.com/en-us/sql/sql-server/sql-server-technical-documentation>.
- [21] MySQL. Mysql documentation. URL: <https://dev.mysql.com/doc/>.
- [22] P. Peng, Y. Tian, Y. Wang, J. Li, and T. Huang. Robust multiple cameras pedestrian detection with multi-view bayesian network. *Pattern Recognition*, 48(5):1760–1772, 2015.
- [23] J. Pokorny. Nosql databases: a step to database scalability in web environment. *International Journal of Web Information Systems*, 9(1):69–82, 2013.
- [24] PostgreSQL. Postgresql documentation. URL: <https://www.postgresql.org/docs/>.
- [25] J. Shang, Z. Wang, and Q. Huang. A robust algorithm for joint sparse recovery in presence of impulsive noise. *IEEE Signal Processing Letters*, 22(8):1166–1170, 2015.
- [26] В.И. Городецкий. Алгебраические байесовские сети—новая парадигма экспертных систем. *Юбилейный сборник трудов институтов Отделения информатики, вычислительной техники и автоматизации РАН*, 2:120–141, 1993.
- [27] В.И. Городецкий and А.Л. Тулупьев. Формирование непротиворечивых баз знаний с неопределенностью. *Известия РАН. Сер. Теория и системы управления*, (5):33–42, 1997.

- [28] А.А. Золотин, Д.Г. Левенец, М.А. Зотов, А.И. Бирилло, А.И. Березин, А.В. Иванова, and А.Л. Тулупьев. Алгоритмы обработки и визуализации алгебраических байесовских сетей. *Образовательные технологии и общество*, 20(1), 2017.
- [29] А.А. Золотин, А.Л. Тулупьев, and А.В. Сироткин. Матрично-векторные алгоритмы локального апостериорного вывода в алгебраических байесовских сетях над пропозициями-квантами. *Научно-технический вестник информационных технологий, механики и оптики*, 15(4), 2015.
- [30] Е.А. Мальчевская and А.А. Золотин. Логико-вероятностный вывод в АБС: архитектура и примеры использования программного комплекса на языке *c#*. In *Гибридные и синергетические интеллектуальные системы*, pages 181–187, 2016.
- [31] А.В. Суворова, А.Л. Тулупьев, and А.В. Сироткин. Байесовские сети доверия в задачах оценивания интенсивности рискованного поведения. *Нечеткие системы и мягкие вычисления*, 9(2):115, 2014.
- [32] А.Л. Тулупьев. Ациклические алгебраические байесовские сети: логико-вероятностный вывод. *Нечеткие системы и мягкие вычисления: Научный журнал Российской ассоциации нечетких систем и мягких вычислений*, 1(1):57–93, 2006.
- [33] А.Л. Тулупьев. *Алгебраические байесовские сети: логико-вероятностная графическая модель баз фрагментов знаний с неопределенностью*. Учреждение Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН, 2009.
- [34] А.Л. Тулупьев, С.И. Николенко, and А.В. Сироткин. Байесовские сети: логико-вероятностный подход, 2006.
- [35] А.Л. Тулупьев, А.В. Сироткин, and А.А. Золотин. Матричные уравнения нормирующих множителей в локальном апостериорном

выводе оценок истинности в алгебраических байесовских сетях. *Вестник Санкт-Петербургского университета. Серия 1. Математика. Механика. Астрономия*, 2(3), 2015.

- [36] А.Л. Тулупьев, А.В. Сироткин, and С.И. Николенко. Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. *СПб.: Изд-во Санкт-Петербургского университета*, 2009.